Queensland University of Technology
Brisbane Australia

This is the author's version of a work that was submitted/accepted for publication in the following source:

Ryan, Caitriona M., Drovandi, Christopher C., & Pettitt, Anthony N.
(2015)
Optimal Bayesian experimental design for models with intractable likelihoods using indirect inference applied to biological process models.
Bayesian Analysis. (In Press)

This file was downloaded from: http://eprints.qut.edu.au/73409/

© Copyright 2015 International Society for Bayesian Analysis

Notice: *Changes introduced as a result of publishing processes such as copy-editing and formatting may not be reflected in this document. For a definitive version of this work, please refer to the published source:*

http://doi.org/10.1214/15-BA977

Bayesian Experimental Design for Models with Intractable Likelihoods using Indirect Inference

C. M. Ryan, C. C. Drovandi and A. N. Pettitt
School of Mathematical Sciences,
Queensland University of Technology, Brisbane, Queensland, Australia
email: 10.ryan@qut.edu.au

July 15, 2014

Abstract

We present indirect inference as a useful method for Bayesian optimal experimental design for models with intractable likelihoods. The approach estimates the parameters of an auxiliary model with a tractable likelihood in an offline step, using simulations from the generative model, the assumed true model of interest. The resulting look-up table is used in the design algorithm of Müller (1999), which uses Markov chain Monte Carlo (MCMC) to sample from an augmented target distribution that is proportional to the expected utility surface. The utility function is based on the precision of a particle approximation of the indirect inference posterior. A novel approach is to consider the gain in utility from the data over and above the utility derived from the prior distribution. The current state-of-the-art method to optimise experimental design for models with intractable likelihoods is to use approximate Bayesian computation (ABC) (Drovandi & Pettitt, 2013). The proposed approach using indirect inference is more flexible, since it is free of the restriction to a discrete design space as in the ABC approach. Moreover, it can be extended to more complex design problems, which would be difficult in the ABC approach, due to the necessity to store all simulations. In contrast, the storage requirements of the indirect inference approach are independent of both the dimension of the experimental design and the number of observations to design for. The use of indirect inference for Bayesian experimental design is illustrated using two stochastic models; a simple death process and an epidemiological model for the evolution of macroparasites, which motivates this work.

Keywords: approximate Bayesian computation; Bayesian experimental design; indirect inference; Markov chain Monte Carlo.

1 Introduction

Experimental design has been fundamental to a wide range of biologically based research activities from seminal work in agriculture (Fisher, 1926), through to recent advances in systems biology (Kreutz & Timmer, 2009), epidemic processes (Cook, et al., 2008), medical science (Delzell et al., 2012) and clinical trials (Berry, 2004). Aspects of data collection are typically limited by the availability of resources such as time, financial cost and sample size. Optimal experimental design aids statistical inference about the underlying process that generates the observed data with the least experimental cost. Factors involved in the design of scientific experiments include the number of experimental units required, the allocation of subjects to treatments and the allocation of available resources.

A common approach is to specify a utility function to reflect the goals and limitations of the experiment. The most useful design should correspond to the design that optimises the utility and is often problem dependent. For example, a design that is optimal for model choice may not be optimal for parameter estimation. In Bayesian statistics, the utility function is typically based on mutual information, that is, the expected Kullback-Leibler divergence (Kullback & Leibler, 1951) from the prior to the posterior distribution.

The focus of this paper is on stochastic models with intractable likelihoods. The aim is to develop Bayesian experimental design methodology for such models in the presence of parameter uncertainty. The motivating example is a stochastic epidemiological model for the evolution of macroparasites. Macroparasites, typically transmitted by mosquitoes, cause lymphatic filariasis disease in an estimated 120 million people worldwide (Ottesen, 2006). A stochastic Markov model for the evolution of *Brugia pahangi* larvae in cats is examined (Michael et al., 1998; Riley et al., 2003). The original experiment (Denham et al., 1972) involved the injection of the larvae in host cats. At various times the cats were sacrificed and the number of mature parasites were recorded. A similar experiment was carried out using mice (Paciorkowski et al., 2000). In a more recent article, Fox et al. (2013) examine the importance of host behaviour and immunity in modelling parasite transmission in a grazing system. For experimental design, the focus is the optimum time to observe the stochastic process for efficient estimation of the model parameters. That is, when to sacrifice the animal.

Much of the experimental design literature, in both the classical and Bayesian paradigms, requires likelihood evaluations. Some exceptions include a recent example to design for Markov process models by Cook et al. (2008), which uses a moment closure approximation to the likelihood to make the utility tractable. The approach is to make a distributional assumption about the state of the process at time t and to close the system by constraining higher-order moments as a function of lower-order moments based on this assumption (Krishnarajah et al., 2005). Although the moment closure approach worked well in the examples presented by Cook et al. (2008), the moment closure approximation will not always adequately represent the behaviour of a stochastic process. This is demonstrated by Riley et al. (2003), where simulation based parameter estimation techniques performed better than moment closure approximations. As a result, the moment closure approach is not useful in the case of the macroparasite model.

The design algorithm of Müller (1999) uses Markov chain Monte Carlo (MCMC) to sample from an augmented target distribution, which has a margin that is proportional to the expected utility surface. The modal utility can be found with its corresponding design. The design algorithm of Müller (1999) is particularly useful in the case of models with intractable likelihoods, due to the cancellation of the likelihood terms in the Metropolis-Hastings acceptance ratio for the design proposal. Drovandi & Pettitt (2013) use an approximate Bayesian computation (ABC) approach to experimental design for models with intractable likelihoods. The utility is specified as a function of the simulation based ABC posterior. This approach is practical since the models are easy to simulate from, but suffers from the usual disadvantages of ABC. A suitable tolerance must be chosen and this will increase as the number of experimental observations to design for increases. For a small number of experimental observations, it is possible to match the simulated data to the observed data directly. This requires storage of a large number of simulations. As a result, the approach does not extend well to complex problems such as designs with a large number of experimental observations to design for or high dimensional designs, where there is more than one controllable variable. Alternatively, summary statistics could be chosen. However, the appropriate choice of summary statistics is not obvious and a loss of information would be incurred by not using the full data. Nevertheless, the approach works well and avoids likelihood evaluations for low dimensional designs with a small number of observations.

In indirect inference, a statistical model is used as an auxiliary for estimation and inference about the generative model. The auxiliary model is based only on observable data and has a tractable likelihood, whereas the generative model may contain unobservables and has an intractable likelihood. Indirect inference has been widely considered as an inferential approach in the frequentist literature (Gourieroux, et al., 1993; Smith, 1993; Gallant & Tauchen, 1996). Some recent advances have been made in the Bayesian literature, combining classical ideas with ABC, where the summary statistics are obtained using the parametric auxiliary model. Drovandi, et al. (2014) present a review of the Bayesian indirect inference methodology.

The use of indirect inference in this paper is similar to Gallant & McCulloch (2009) and Reeves & Pettitt (2005). Instead of using the auxiliary parameter estimates as summary statistics in ABC, we do not reduce the data to a set of summary statistics. Rather, the auxiliary likelihood of simulated data from the generative model is evaluated at the auxiliary parameter estimates. This parametric likelihood is used in the analysis as a substitute for the intractable likelihood of the generative model. Drovandi et al. (2014) refer to this method as pdBIL (parametric Bayesian indirect likelihood on the full data level) and demonstrate that this method is not exact and will rarely target the true posterior. However, Drovandi et al. (2014) suggest that, if the auxiliary model likelihood acts as a good replacement to the generative model likelihood for non-negligible regions of the posterior, then a useful approximation can be obtained. Gallant & McCulloch (2009) apply this approach to modelling asset pricing.

In this paper, we present a novel approach to Bayesian experimental design for models with intractable likelihoods using indirect inference and demonstrate its effectiveness in the

case of Bayesian experimental design for stochastic biological processes. The utility function is specified as a function of an indirect inference posterior, which is based on a tractable auxiliary model. Having found a suitable auxiliary model, the approach is to learn the relationship between the auxiliary model parameters and the data model parameters using simulations from the data model and maximum likelihood estimation in an offline step. Auxiliary parameter values and their corresponding generative model parameter draws from the prior are stored and used in the design algorithm of Müller (1999) for Bayesian design optimisation. An important advantage of the indirect inference approach to experimental design is that the storage requirements are independent of the number of experimental observations and the dimension of the design. Furthermore, the ABC approach of Drovandi & Pettitt (2013) must be carried out over a discrete design space, whereas the indirect inference approach is free of this restriction.

The paper describes the Bayesian approach to optimal experimental design in Section 2. Section 3 explains the novel indirect inference approach to Bayesian optimal design. Section 4 illustrates the methodology using the death model and the stochastic epidemiological model for the evolution of macroparasites. The paper concludes with a discussion in Section 5.

2 Bayesian experimental design

This paper concerns the Bayesian design of experiments, where the design of interest is the choice of the sampling times of a stochastic process, based on prior beliefs. The prior distribution $p(\boldsymbol{\theta})$, can be elicited from experts or based on previous experiments. The aim is to choose observation times that are expected to lead to the most precise posterior distribution. Following Lindley (1972), the design choice is framed as a decision problem by specifying a suitable utility function $u(d, \mathbf{y}, \boldsymbol{\theta})$, where $\mathbf{y}$ are the data that may be observed when experimental design d is applied and where $\boldsymbol{\theta}$ is the vector of model parameters. The Bayesian optimal design corresponds to the maximum posterior expected utility. The problem of Bayesian experimental design can be addressed by the following optimisation for designs $d \in D$;

$$d^* = \underset{d \in D}{\operatorname{argmax}}\ u(d), \tag{1}$$

where $u(d)$ is the expected utility function over the data $\mathbf{y}$ and the prior distribution of the model parameters $\boldsymbol{\theta}$;

$$u(d) = \mathbf{E}_{\boldsymbol{\theta},\mathbf{y}}[u(d, \mathbf{y}, \boldsymbol{\theta})] = \int_{\mathbf{y}} \int_{\boldsymbol{\theta}} u(d, \mathbf{y}, \boldsymbol{\theta}) p(\mathbf{y}|\boldsymbol{\theta}, d) p(\boldsymbol{\theta}) d\boldsymbol{\theta} d\mathbf{y}. \tag{2}$$

However, integration across all possible values of the data $\mathbf{y}$ that are yet to be observed and model parameters $\boldsymbol{\theta}$ is typically intractable. In our case the likelihood $p(\mathbf{y}|\boldsymbol{\theta}, d)$ given the design d is also intractable.

A pragmatic approach taken by Müller (1999) is to sample from an augmented joint probability distribution;

$$h(d, \boldsymbol{\theta}, \mathbf{y}) \propto u(d, \mathbf{y}, \boldsymbol{\theta}) p(\mathbf{y}|\boldsymbol{\theta}, d) p(\boldsymbol{\theta}). \tag{3}$$

Note that the normalising constant required for this distribution for fixed d is the expected utility $u(d)$ of Equation 2. Thus the margin over $\mathbf{y}$ and $\boldsymbol{\theta}$ of the target distribution $h(d, \boldsymbol{\theta}, \mathbf{y})$ is proportional to $u(d)$ and shares the same mode. Müller (1999) borrows ideas from simulated annealing (Van Laarhoven & Aarts, 1987), tempering the distribution by a single temperature J. This exaggerates the peaks of the distribution and enables an easier search for the mode. The tempered target distribution is

$$h_J(d, \boldsymbol{\theta}, \mathbf{y}) \propto \prod_{j=1}^{J} u(d, \mathbf{y}_j, \boldsymbol{\theta}_j) p(\mathbf{y}_j | d, \boldsymbol{\theta}_j) p(\boldsymbol{\theta}_j), \tag{4}$$

where $\boldsymbol{\theta}_j$ is an independent draw from the prior distribution of the generative model parameters and $\mathbf{y}_j$ is a simulation from the generative model given the design d and $\boldsymbol{\theta}_j$. Increasing the value of J tightens the distribution further at its mode, but incurs a computational cost. Algorithm 1 outlines the design algorithm of Müller (1999). In line 4, Algorithm 1, $q(.|.)$ denotes the Metropolis-Hastings proposal distribution. The algorithm is particularly appealing for models with intractable likelihoods due to the cancellation of the likelihood terms from the Metropolis-Hastings ratio. This results from the simulation of the proposed data $\mathbf{y}^*$ from the likelihood in line 5, Algorithm 1. However, the utility function $u(d, \mathbf{y}, \boldsymbol{\theta})$ typically requires likelihood evaluations. The next section demonstrates how the utility can be approximated without calculating the likelihood of the generative model, using indirect inference.

Algorithm 1: MCMC algorithm for robust Bayesian experimental design (Müller, 1999).

INPUT: Initial design d^0, number of iterations T;

1 Draw $\boldsymbol{\theta}_j^0 \sim p(\boldsymbol{\theta})$ and simulate $\mathbf{y}_j^0 \sim p(\mathbf{y}|\boldsymbol{\theta}_j^0, d^0)$ for $j = 1 \dots J$;

2 Calculate the utility $u_j^0 = u(d^0, \boldsymbol{\theta}_j, \mathbf{y}_j^0)$ for $j = 1 \dots J$ and define $u^0 = \prod_{j=1}^{J} u_j^0$;

3 **for** $t = 1, \dots, T$ **do**

4 Draw $d^* \sim q(d^*|d^{i-1})$;

5 Draw $\boldsymbol{\theta}_j^* \sim p(\boldsymbol{\theta})$ and simulate $\mathbf{y}_j^* \sim p(\mathbf{y}|\boldsymbol{\theta}_j^*, d^*)$ for $j = 1 \dots J$;

6 Calculate $u_j^* = u(d^*, \boldsymbol{\theta}_j^*, \mathbf{y}_j^*)$ for $j = 1 \dots J$ and let $u^* = \prod_{j=1}^{J} u_j^*$;

7 Set $(d^t, u^t) = (d^*, u^*)$ with probability $\min(1, \alpha)$ where,

$$\alpha = \frac{u^*}{u^{t-1}} \frac{q(d^{t-1}|d^*)}{q(d^*|d^{t-1})}.$$

Otherwise set $(d^t, u^t) = (d^{t-1}, u^{t-1})$;

8 **end**

OUTPUT: Marginal draws of the designs d;

3 Bayesian experimental design using indirect inference

Assume there exists a statistical model, which is a good approximation to the generative model in regions of high posterior density. The parameters of the auxiliary model are denoted $\boldsymbol{\phi}$ and for a given design d, the tractable auxiliary likelihood is denoted $p_a(\mathbf{y}|\boldsymbol{\phi}, d)$, where the subscript a denotes the auxiliary model. The parameters of the generative model are denoted $\boldsymbol{\theta}$ and its likelihood $p(\mathbf{y}|\boldsymbol{\theta}, d)$ given design d is intractable.

For Bayesian experimental design using indirect inference, the approach is to learn the relationship between the auxiliary model parameters and the data model parameters using a training design d_T in an offline step via Algorithm 2. The training design could be of the same structure as in previous experiments, but the user is free to choose this design (more information is provided in the discussion in Section 5). For $\boldsymbol{\theta}^i$, the ith value of $\boldsymbol{\theta}$ generated from the prior, a dataset $\mathbf{x}$ is simulated from the generative model based on the training design d_T. The auxiliary parameters are estimated given this data, producing $\boldsymbol{\phi}^i = \boldsymbol{\phi}(\boldsymbol{\theta}^i, \mathbf{x})$, where

$$\boldsymbol{\phi}^i = \underset{\boldsymbol{\phi} \in \Phi}{\operatorname{argmax}}\ p_a(\mathbf{x}|\boldsymbol{\phi}, d_T). \tag{5}$$

Repeating this process for a collection of n parameters from the prior produces a look-up table $\{\boldsymbol{\theta}^i, \boldsymbol{\phi}^i\}_{i=1}^n$. This process effectively produces a noisy estimate $\boldsymbol{\phi}(\boldsymbol{\theta}, \mathbf{x})$ of the mapping $\boldsymbol{\phi}(\boldsymbol{\theta})$, between the generative and auxiliary parameters. In an attempt to reduce the noise, one can instead simulate m independent replicates of the data, denoted $\mathbf{x}_{1:m} = (\mathbf{x}_1, \ldots, \mathbf{x}_m)$. The corresponding $\boldsymbol{\phi}^i$ is given by $\boldsymbol{\phi}^i = \boldsymbol{\phi}(\boldsymbol{\theta}^i, \mathbf{x}_{1:m})$. Under the assumption that the auxiliary estimator is consistent, the true mapping $\boldsymbol{\phi}(\boldsymbol{\theta})$, for a particular value of $\boldsymbol{\theta}$, can be recovered as $m \to \infty$. Increasing the value of m results in a more precise determination of the mapping for a particular value of $\boldsymbol{\theta}$ while increasing n allows the mapping to be approximated for more points across the prior space. For the purposes of auxiliary parameter estimation we recommend maximum likelihood estimation in line 4, Algorithm 2, since it generally leads to more statistically efficient estimators compared with other techniques, but we note that other methods could be used such as the method of moments. It is important to stress that with the indirect inference approach we only need to store the look-up table $\{\boldsymbol{\theta}^i, \boldsymbol{\phi}^i\}_{i=1}^n$, whose size is independent of the number of experimental observations and the number of design dimensions. In contrast, the ABC approach of Drovandi & Pettitt (2013) requires storage of all the simulated data, whose size grows exponentially with the number of design dimensions. Furthermore, as the number of experimental observations grows, the ABC tolerance will increase. This can be mitigated by increasing the number of prior simulations but this again increases the storage requirements. Having estimated the map from the generative model to the auxiliary statistical model, indirect inference can be used to optimise Bayesian experimental design. The design algorithm of Müller (1999), Algorithm 1, is employed, where the utility is a function of the indirect inference posterior in place of the true posterior. This is similar to the approach of Drovandi & Pettitt (2013), where the utility is a function of the ABC approximation to the true posterior. The choice of utility function for this problem was

Algorithm 2: Offline step to estimate $\boldsymbol{\phi}(\boldsymbol{\theta})$.

INPUT: Training design d_T, number of simulations n;

for $i = 1, \ldots, n$ **do**

  Draw $\boldsymbol{\theta}^i \sim p(\boldsymbol{\theta})$;

  Simulate $\mathbf{x}_j \sim p(\mathbf{y}|\boldsymbol{\theta}^i, d_T), \quad j = 1, \ldots, m$;

  Find $\boldsymbol{\phi}^i = \underset{\boldsymbol{\phi}\in\Phi}{\operatorname{argmax}} \prod_{j=1}^{m} p_a(\mathbf{x}_j|\boldsymbol{\phi}, d_T)$;

end

OUTPUT: Look-up table $\{\boldsymbol{\theta}^i, \boldsymbol{\phi}^i\}_{i=1}^n$;

driven by the goal of parameter estimation. One useful utility is the 'Bayesian D-posterior precision' (Drovandi et al., 2013), which is a function of the precision of the indirect inference posterior distribution $p_a(\boldsymbol{\theta}|\mathbf{y}, d, \boldsymbol{\phi}(\boldsymbol{\theta}, \mathbf{x}))$, defined by

$$u(d, \mathbf{y}) = 1/\det(\widehat{\mathrm{Var}}(\boldsymbol{\theta}|\mathbf{y}, d, \boldsymbol{\phi}(\boldsymbol{\theta}, \mathbf{x}))). \tag{6}$$

This is appropriate for unimodal posterior distributions that are not significantly skewed. However, we found that the use of this utility was relatively flat across the design space. In order to sharpen the utility surface, a large value of $J > 100$ is required in line 6 of Algorithm 1, which is highly computationally intensive. Instead we propose the following utility, which represents a gain in utility from the data over and above the utility found from the informative prior;

$$\tilde{u}(d, \mathbf{y}) = \max\{(u(d, \mathbf{y}) - su_p), 0\}, \tag{7}$$

where u_p is the utility calculated from the prior distribution;

$$u_p = 1/\det(\mathrm{Var}(\boldsymbol{\theta})), \tag{8}$$

and where $s \leq 1$ is a scaling factor on u_p to aid mixing of the Markov chain in Algorithm 1. This is a similar idea to the Kullback-Leibler divergence (Kullback & Leibler, 1951) from the prior to the posterior, which is often used as a utility in Bayesian experimental design (Cook et al., 2008; Drovandi et al., 2013; Huan & Marzouk, 2012) but is unavailable in the case of models with intractable likelihoods. Using the utility of Equation 7, we found that the utility surface was substantially sharpened using a value of only 10 for J. The idea is demonstrated in Figure 1(a) for one observation of the macroparasite model. The mean utility of $1,000$ simulations is plotted for 15 discrete observation times. This is possible for $k = 1$ observation but difficult for $k > 1$ due to the exponential growth of the discretisation of the design with the number of observations k. The solid curve is a *lowess* smooth (Cleveland, 1979) of the calculated mean utilities. The dashed line in Figure 1(a) is the prior utility, u_p. This bound can be lowered using s in Equation 7 to ensure that the utility remains positive with probability close to 1. Note that the y-axis in Figure1(a) is truncated below at 4.4×10^{11} and that the utility $u(d, \mathbf{y})$ relative to the origin is a very flat surface.

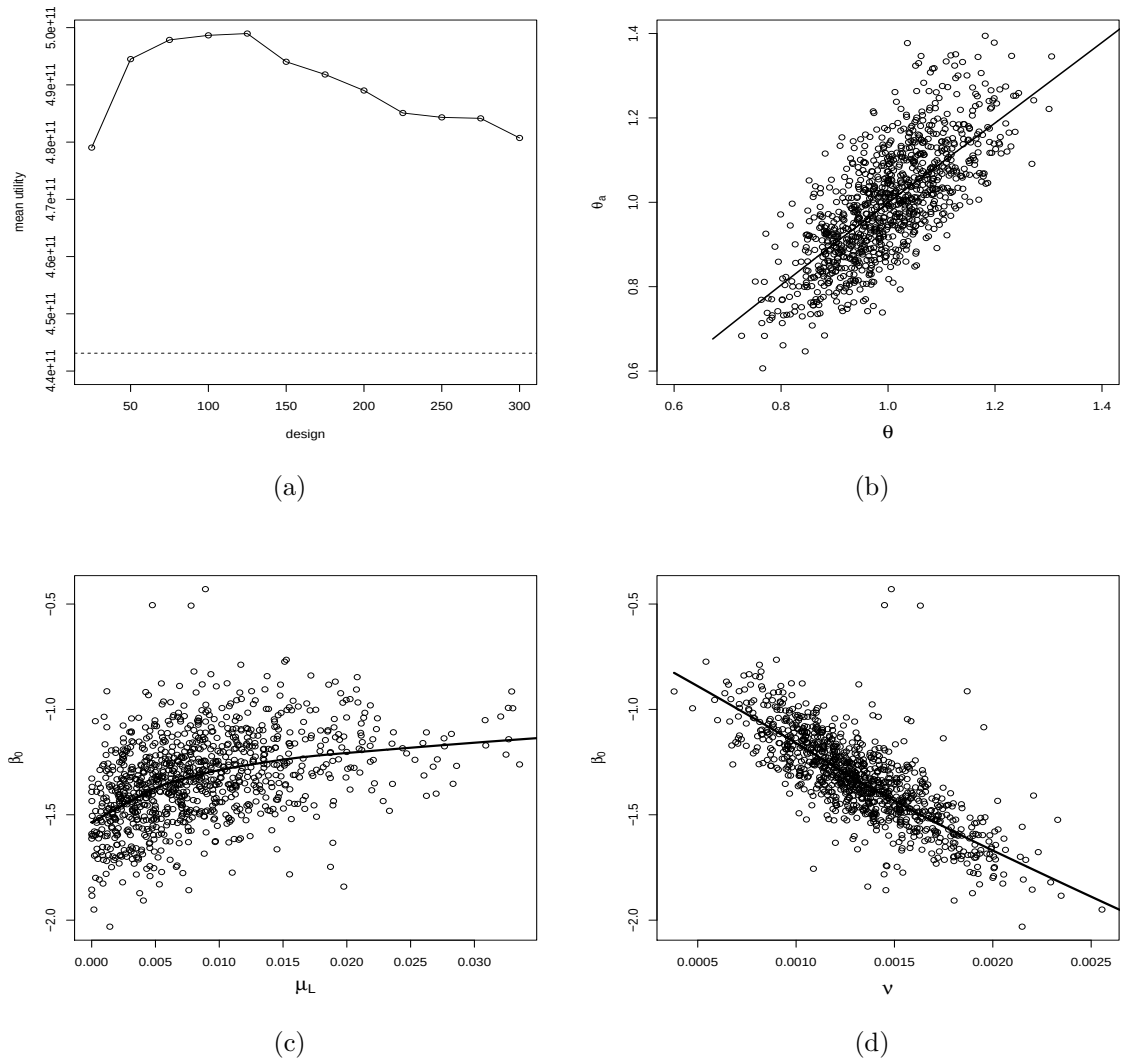

Figure 1: Plot (a) displays the mean utility of Equation 6 for 1 observation calculated using $1,000$ simulations over a fixed grid of designs, where the dashed line is the prior utility u_p and the solid curve is a *lowess* smooth (Cleveland, 1979) of the mean utilities. Plot (b) displays the marginal relationship between the auxiliary parameter θ_a and the generative model parameter θ for the death model. Plots (c) and (d) respectively display the marginal relationship between the auxiliary parameter β_0 and generative model parameters $\boldsymbol{\theta} = (\mu_L, \nu)$ for the macroparasite example, having carried out the offline Algorithm 2. In plots (b) to (d), the black line is a *lowess* smooth (Cleveland, 1979), to indicate the relationship between these generative and auxiliary parameters.

The utility in Equation 6 is calculated using Algorithm 3, where an indirect inference posterior for $\boldsymbol{\theta}$ is formed using importance sampling, where the prior $p(\boldsymbol{\theta})$ is the importance distribution and the ith importance weight is proportional to $p_a(\mathbf{y}|\boldsymbol{\phi}^i, d)$ for $i = 1, \ldots, n$. This is deterministic given the look-up table $\{\boldsymbol{\theta}^i, \boldsymbol{\phi}^i\}_{i=1}^n$ that is precomputed in Algorithm 2. This ensures fully reversible Metropolis-Hastings moves in Algorithm 1, satisfying detailed balance. Otherwise a pseudo-marginal approach would be necessary (Andrieu & Roberts, 2009). This is similar to the approach of Huan & Marzouk (2012), who use a sample average approximation combined with a deterministic quasi-Newton method to optimise the design of experiments for non-linear systems. As in our case, the same set of "noise" random variables is used for different values of d, which makes the design optimisation problem deterministic, rather than stochastic.

Algorithm 3: Calculation of the utility $u(d, \mathbf{y})$ using a particle approximation of the indirect inference posterior distribution.

INPUT: Simulated data $\mathbf{y}$, look-up table $\{\boldsymbol{\theta}^i, \boldsymbol{\phi}^i\}_{i=1}^n$, design d;

1 Calculate importance weights $W_i \propto p_a(\mathbf{y}|\boldsymbol{\phi}^i, d)$ for $i = 1, \ldots, n$;

2 Set $u(d, \mathbf{y}) = 1/\det(\widehat{\mathrm{Var}}(\boldsymbol{\theta}|\mathbf{y}, d, \boldsymbol{\phi}(\boldsymbol{\theta}, \mathbf{x})))$, where $\mathrm{Var}(\boldsymbol{\theta}|\mathbf{y}, d, \boldsymbol{\phi}(\boldsymbol{\theta}, \mathbf{x}))$ is estimated using a particle approximation $\{W_i, \boldsymbol{\theta}^i\}_{i=1}^n$ of the indirect inference posterior distribution;

OUTPUT: Utility $u(d, \mathbf{y})$;

Having learned the noisy mapping $\boldsymbol{\phi}(\boldsymbol{\theta}, \mathbf{x})$ using Algorithm 2, the resulting look-up table $\{\boldsymbol{\theta}^i, \boldsymbol{\phi}^i\}_{i=1}^n$ and the utility function of Equation 6 can be used in the design algorithm of Müller (1999), Algorithm 1, to sample from the following tempered target distribution;

$$h_J(d, \boldsymbol{\theta}, \mathbf{y}) \propto \prod_{j=1}^{J} u(d, \mathbf{y}_j) p(\mathbf{y}_j|\boldsymbol{\theta}_j, d) p(\boldsymbol{\theta}_j). \tag{9}$$

The use of the indirect inference posterior distribution in the utility calculation allows the algorithm to be carried out without the calculation of the intractable likelihood of the generative model. The approach is illustrated using the simple death process and the macroparasite model in the following section.

4 Results

Bayesian experimental design using indirect inference is demonstrated using two examples. The first is the death model (Renshaw, 1991), which is used as an illustrative problem since the likelihood for the model is easy to compute. Comparisons can be made between the design algorithm of Müller (1999) using indirect inference to the same algorithm using the true likelihood as well as with the approach to Bayesian experimental design using the moment closure approximation to the likelihood (Cook et al., 2008). The second example motivates this work and concerns the population evolution of macroparasites (Michael et al.,

1998; Riley et al., 2003). The likelihood is computationally intractable. This necessitates the use of approximate methods to perform inference for the experimental design. The experimental design of interest is the optimal time to observe the stochastic process. That is, the time to sacrifice the animal. It is straightforward to simulate from this model using the algorithm of Gillespie (1977).

Combinations of values of $n \leq 100,000$ and $m = (3, 6, 10)$ in Algorithm 2 were examined in our examples. $m = 3$ and $n = 10,000$ worked well for both the death model and the motivating macroparasite model. Values of $m \leq 10$ did not have much effect on the auxiliary parameter estimates. Assuming that the auxiliary estimator is consistent, as $m \to \infty$, $\boldsymbol{\phi}(\boldsymbol{\theta}, \mathbf{x}) \to \boldsymbol{\phi}(\boldsymbol{\theta})$, reducing the noise in the auxiliary parameter estimates. However, one might expect that values of m close to 100 or more would be required to have an effect. In an application of the indirect inference approach to modelling asset pricing, Gallant & McCulloch (2009) use $m \approx 700$ and an MCMC algorithm to find the maximum likelihood estimator. This increase in computational complexity is deemed unnecessary in the context of optimal experimental design. Our approach is to use $m = 3$ and in line 4 of Algorithm 2, Nelder-Mead optimisation (Nelder & Mead, 1965), which is more pragmatic and works well for our examples.

The target distribution was tempered using $J = 10$ in line 5, Algorithm 1, which was deemed sufficient for our examples using the utility of Equation 7 with $s \approx 0.8$. Increasing J and s exaggerates the peaks of the marginal densities of the designs. However, increasing J considerably increases the computation time. Large values of J or long runs of the design algorithm of Müller (1999) are required for low values of s. For example, in the case of one observation, the use of $s = 0$, requires $J = 50$ to achieve similar results in the same number of iterations of the design algorithm of Müller (1999) as the use of $s = 0.8$ with $J = 10$. One might expect that a much larger value of J such as 100 or more may have a stronger effect on the densities. However, this would be highly computationally intensive and is deemed unnecessary in the context of our optimal experimental design problems.

4.1 Death model example

4.1.1 Generative model

The simple death process (Renshaw, 1991) is used to illustrate the methodology. At time t, with $S(t) = i$ susceptibles, the probability that an infection occurs in the next infinitesimal time period Δt is given by

$$p\left(S(t + \Delta t) = i - 1 | S(t) = i\right) = \theta i \Delta t + o(\Delta t).$$

The observable data are susceptible, time pairs $(t_1, S_1(t_1)), \ldots (t_T, S_T(t_T))$, where $S_j(t_j)$ comes from a binomial distribution;

$$S_j(t_j) \sim \text{binomial}\left(S_{j-1}(t_{j-1}), \mathbf{e}^{-\theta(t_j - t_{j-1})}\right), \quad j = 1, \ldots, T. \tag{10}$$

The process is initialised with one infected such that the number of susceptibles at time t_0 is $S_0(0) = n - 1$.

4.1.2 Auxiliary model

To demonstrate the methodology, a normal distribution is used as an auxiliary model;

$$S_j(t_j) \sim \text{normal}\left(S_{j-1}(t_{j-1})\mathbf{e}^{-\theta_a(t_j^\gamma - t_{j-1}^\gamma)}, \sigma^2 S_{j-1}(t_{j-1})\mathbf{e}^{-\theta_a(t_j^\gamma - t_{j-1}^\gamma)}\left(1 - \mathbf{e}^{-\theta_a(t_j^\gamma - t_{j-1}^\gamma)}\right)\right). \tag{11}$$

This allows a more flexible variance and time scale than the binomial distribution. This is a poor approximation for S_{j-1} close to 0, giving support to negative values of S_j. Nevertheless, it is useful for illustrative purposes and performs well for the design problem. For this example, the auxiliary parameters are $\boldsymbol{\phi} = (\theta_a, \gamma, \sigma^2)$ and the generative model parameter is θ. The prior distribution for θ follows Cook et al. (2008), where $\theta \sim \log\text{normal}(-0.005, 0.01)$. Figure 2(a) displays the median of $100,000$ data simulations from the death model prior together with 95% prediction intervals.

4.1.3 Design results

For the indirect inference approach to Bayesian experimental design, Algorithm 2 is carried out offline to estimate the map $\boldsymbol{\phi}(\boldsymbol{\theta})$ via the resulting look-up table $\{\boldsymbol{\theta}^i, \boldsymbol{\phi}^i\}_{i=1}^n$. For the purposes of illustration, the training design d_T is 20 equispaced observations across 10 days and the initial number of susceptibles is 50. The resulting densities of σ^2 and γ were centred close to 1, which was expected since the normal approximation to the binomial distribution for large $S_j(t_j)$, is Equation 11, where $\sigma^2 = \gamma = 1$. A strong relationship between θ_a and θ can be seen in Figure 1(b).

The design algorithm of Müller (1999) was carried out for $100,000$ iterations of each of one to four observation times. Designs d of interest are between $d_{min} = 0$ days and $d_{max} = 10$ days. The algorithm cycles through each design point in turn, given the current value of all other design points. Observations are ordered such that $d_1 < d_2 < \ldots < d_k$. At iteration t, the proposed value of d_i^t is generated from a truncated normal random walk with variance 1, truncated at d_{i-1}^t below and d_{i+1}^{t-1} above for $i = 1, \ldots, k$ (where the notation assumes that $d_0^t = d_{min}$ and $d_{k+1}^{t-1} = d_{max}$). Figure 3 displays the resulting density estimates for 1 to 4 observations (continuous lines). The likelihood of the death model is tractable and the approach was also carried out using the true likelihood in place of the auxiliary likelihood in the utility calculation of Algorithm 3. The resulting marginal densities are similar and are plotted in Figure 3 (dashed lines).

For more than one observation time, the optimal design differs from the modes of the marginal densities plotted in Figure 3, since the designs are time ordered. Consider the case of two ordered observations drawn from uniform densities. The marginal modes would appear at the endpoints 0 and d_{max}. A Gaussian smoothing kernel was used to estimate the modal designs, finding the design that has the most other designs in its vicinity. The resulting optimal designs are displayed in Table 1 using automatic bandwidth estimated by the `KernSmooth` package in `R` via the plug-in method of Wand & Jones (1994). We follow Drovandi & Pettitt (2013), tempering the bandwidth matrix by smoothing factors $h \in (0.1, 0.2, ..., 4)$ and optimising the density at each bandwidth. Possible alternatives

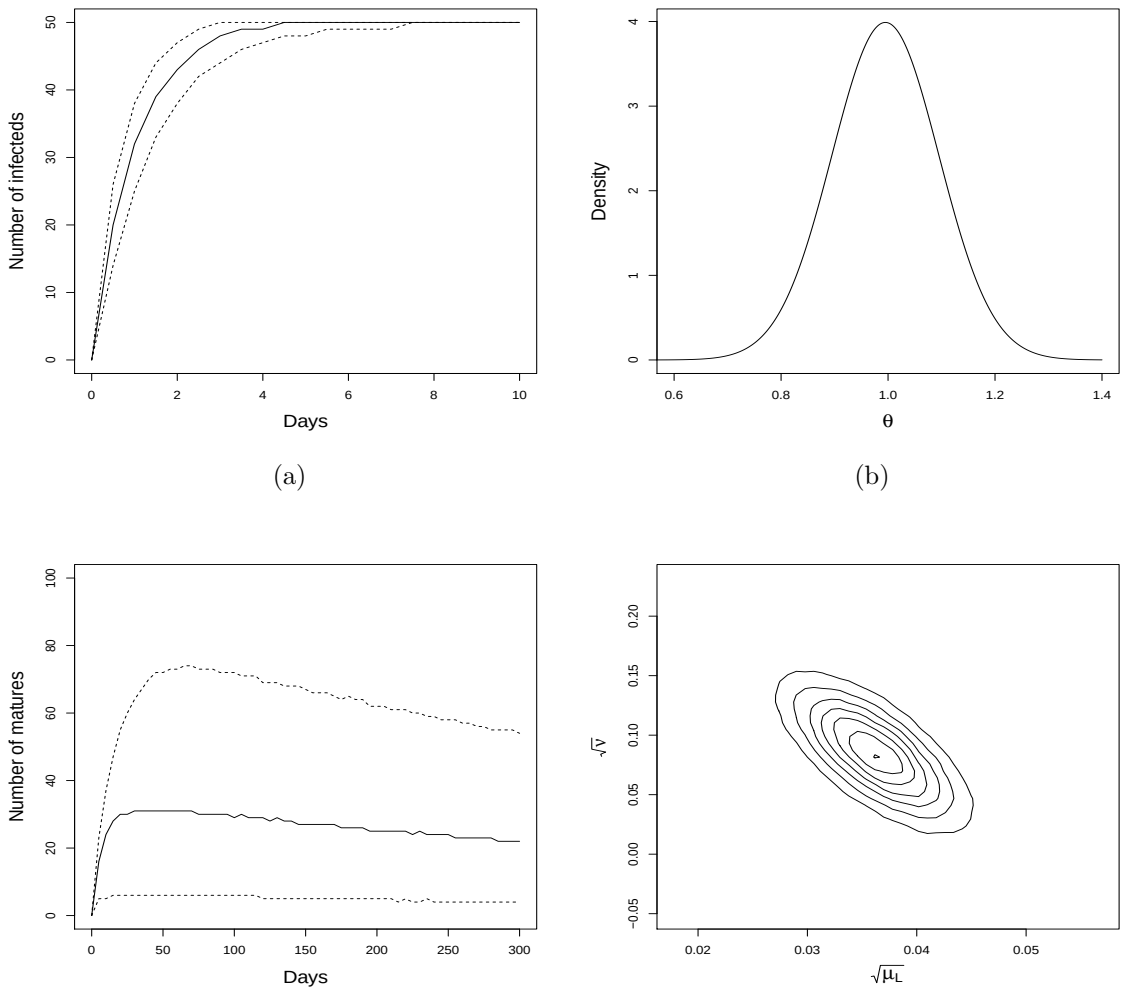

Figure 2: Median (solid line) and 95% prior prediction intervals (dashed lines) of $100,000$ simulated counts of infecteds from the death model (plot (a)) from a log normal$(-0.005, 0.001)$ prior distribution for θ (plot (b)) and median (solid line) and 95% prior prediction intervals (dashed lines) of $100,000$ simulated counts of mature parasites from the macroparasite model (plot (c)) where $l_i = 100$ initial larvae were injected in each host and where $(\sqrt{\nu}, \sqrt{\mu_L})$ come from a bivariate normal distribution a priori (plot (d)).

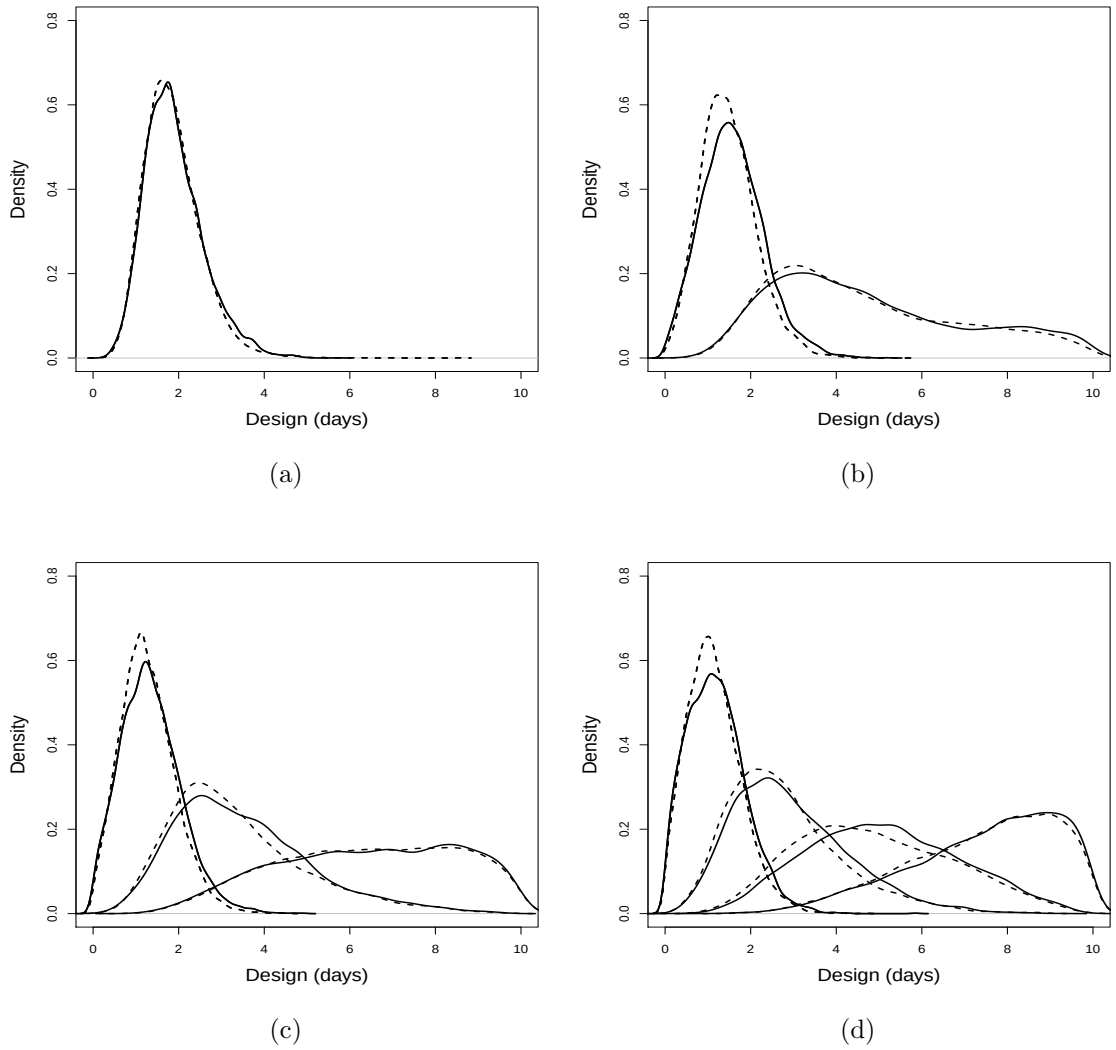

Figure 3: Marginal density estimates of the designs for up to 1 to 4 observations of the death model (plots (a) to (d) respectively) for $100,000$ iterations of the design algorithm of Müller (1999) using indirect inference. The dashed lines are based on $100,000$ iterations of the design algorithm of Müller (1999) using the true binomial likelihood in place of the auxiliary likelihood in the utility calculation of Algorithm 3.

Table 1: Optimal designs for $k = 1, \ldots, 4$ observations of the death model using the design algorithm of Müller (1999) and indirect inference (II) and the moment closure (MC) approximation (results from Cook et al. (2008)). Both approaches use a Gaussian smoothing kernel to find the modal values. Expected utilities, Equation 12, are calculated using indirect inference at the optimal design inferred by each approach.

k	II optimal design d	$u(d)$	MC optimal design d	$u(d)$
1	1.75	130.40	1.7	129.37
2	$(1.2, 3.3)$	135.92	$(0.9, 2.4)$	135.64
3	$(0.8, 2.2, 4.2)$	138.42	$(0.7, 1.5, 2.9)$	138.24
4	$(0.6, 0.9, 2.0, 3.6)$	139.41	$(0.8, 1.7, 3.1, 5.3)$	138.83

to this approach are discussed in Section 5. Having calculated the expected utility of any resulting distinct designs, the optimum design is chosen accordingly. The expected utility is calculated using simulations $\mathbf{y}_j$ from the generative model, where

$$u(d) \approx \sum_{j=1}^{10000} u(d, \mathbf{y}_j), \tag{12}$$

and where $u(d, \mathbf{y}_j)$ is calculated using Algorithm 3. Results are displayed in Table 1 and are similar to the optimal designs inferred by Cook et al. (2008) using the moment closure approach.

4.2 Macroparasite model example

The motivating example for this work is a stochastic Markov model for the evolution of L3 *Brugia pahangi* larvae in cats (Michael et al., 1998; Riley et al., 2003). In the original experiment (Denham et al., 1972), approximately 100 or 200 larvae were injected in 21 host cats that had never previously been exposed to the parasite. Microfilarial counts, that is, the number of mature parasites, were recorded weekly in each cat and again at various autopsy times between 24 and 1,193 days. Measurements also made at that time but not published are included in Michael et al. (1998) and Riley et al. (2003) for 212 cats with precisely recorded initial L3 larvae counts and final mature parasite counts at autopsy time. This data is used to inform the prior distribution $p(\boldsymbol{\theta})$ for future experiments. The training design d_T is based on the same experimental design used to obtain the observed data. The experiments were conducted over a number of years in which the availability of the larvae varied. A similar experiment was recently carried out using mice (Paciorkowski et al., 2000), where 50 L3 larvae were injected and mice were necropsied at fixed time points after 1 to 6 weeks post-infection. The design of interest is the optimum time to observe the stochastic process, that is, the time of sacrifice of the cat, with the goal of parameter estimation.

4.2.1 Generative model

The following stochastic model was developed to explain the within-host population dynamics of lymphatic filariasis (Michael et al., 1998; Riley et al., 2003) . At time t, a host is described by the following random variables, the mature parasite count $M(t)$, the larvae count $L(t)$ and a discrete measure of the host's immunity $I(t)$. The host cats have never previously been exposed to the parasite and thus have no experience of infection, no immunity and no mature parasites. This gives initial conditions of $I(0) = 0$, $M(0) = 0$ and $L(0) = l_i$, since the initial number of L3 larvae l_i for cat i is injected at time 0. The values of the states at time t are $M(t) = i$, $L(t) = j$ and $I(t) = k$. The larvae mature to adult parasites, die due to the natural death of the larvae, or die due to the immune response of the host. Note that the number of larvae is unobservable at autopsy, as is the level of immunity. Only the number of mature parasites is counted. The mature parasites die at a rate μ_M per day. Larvae mature at a rate γ per day. Larvae die at a rate $\mu_L + \beta I(t)$ per day, where μ_L corresponds to the natural death rate of the larvae and β is the additional death rate due to the host's immune response. The host's immune response, $I(t)$, is assumed to increase at a rate ν per larva per day and decrease at a rate μ_I per unit of immunity. The immune response is assumed to affect larvae only and not the mature parasites. Thus for a small time interval Δt such that at most one event can occur, the transition probabilities at time $t + \Delta t$ are given by;

$$\begin{aligned}
p(i+1, j-1, k) &= \gamma j \Delta_t + o(\Delta_t), \\
p(i, j-1, k) &= (\mu_L + \beta k) j \Delta_t + o(\Delta_t), \\
p(i-1, j, k) &= \mu_M i \Delta_t + o(\Delta_t), \\
p(i, j, k+1) &= \nu j \Delta_t + o(\Delta_t), \\
p(i, j, k-1) &= \mu_I k \Delta_t + o(\Delta_t).
\end{aligned} \tag{13}$$

4.2.2 Auxiliary model

The observed data is the initial larvae count l_i and the mature larvae count m_i at autopsy time t_i for hosts $i = 1, \ldots, n$. An auxiliary beta-binomial distribution based on the observed data has been used in the literature (Drovandi et al. (2011)) where,

$$p(M(t_i) = m_i | \alpha_i, \beta_i) = \binom{l_i}{m_i} \frac{B(m_i + \alpha_i, l_i - m_i + \beta_i)}{B(\alpha_i, \beta_i)}.$$

This is re-parameterised in terms of a proportion $p_i = \alpha_i/(\alpha_i + \beta_i)$ and overdispersion $\xi_i = 1/(\alpha_i + \beta_i)$. The auxiliary model relates these parameters to the covariates (t_i, l_i) via;

$$\begin{aligned}
\text{logit}(p_i) &= f_p(t_i, l_i) = \beta_0 + \beta_1(\log(t_i) - \overline{\log(t)}) + \beta_2(\log(t_i) - \overline{\log(t)})^2. \\
\log(\xi_i) &= f_\xi(t_i, l_i) = f_\xi(l_i) = \begin{cases} \eta_1, & \text{if } l_i \leq 100 \\ \eta_2, & \text{otherwise} \end{cases}.
\end{aligned}$$

With the above re-parameterisation, the auxiliary parameters are $\boldsymbol{\phi} = (\beta_0, \beta_1, \beta_2, \eta_1, \eta_2)$ and the generative model parameters are $\boldsymbol{\theta} = (\nu, \mu_I, \mu_L, \beta)$.

4.2.3 Design results

The model parameters used for the design of experiments using indirect inference are (ν, μ_L) as they can be reasonably well estimated by mature parasite counts (Drovandi et al., 2011) and they relate to the auxiliary parameter estimates $\boldsymbol{\phi}$ consistently across the parameter space of interest. The prior distributions are taken from a previously published study (Drovandi et al., 2011), where $(\sqrt{\nu}, \sqrt{\mu_L})$ are approximated by a bivariate normal distribution with mean $(0.0361, 0.0854)$ and standard deviations $(0.0045, 0.0342)$ with a correlation of -0.6974. Figure 2(c) displays the median of $100,000$ data simulations from this prior together with 95% prediction intervals. Other parameters are assumed known at estimates found in previous studies; $\gamma = 0.04$ (Suswillo, et al., 1982), $\mu_M = 0.0015$ (Michael et al., 1998), $\beta = 1.1$ and $\mu_I = 0.31$ (Riley et al., 2003). Note that β and μ_I can be difficult to estimate from observed mature counts only (Drovandi et al., 2011). To make comparisons with Drovandi & Pettitt (2013), designs are restricted to earlier than $d_{max} = 300$ days. Designs are also restricted to be later than $d_{min} = 24$ days, as this is the minimum autopsy time of the original training design d_T. Beyond approximately 75 days there are no juveniles left a priori, after which, the mature parasites follow a simple death process with death rate μ_M, since the immunity is affected only by the juvenile larvae. Note that the number of juveniles is unobserved at autopsy time and only the number of mature parasites is recorded.

The offline step, Algorithm 2, was carried out for $m = 3$ and $n = 10,000$. Figure 1(c) and Figure 1(d) display the marginal relationship between $\boldsymbol{\theta} = (\mu_L, \nu)$ and β_0 as an example of the relationship between the generative model parameters $\boldsymbol{\theta}$ and the auxiliary parameters $\boldsymbol{\phi}$. The black line is a *lowess* smooth (Cleveland, 1979), which indicates a strong relationship between these generative and auxiliary parameters. Results for $100,000$ iterations of the design algorithm of Müller (1999) are displayed in Figure 4 for the macroparasite exaumple. The proposal distribution is a truncated normal random walk as described in the death model example, Section 4.1.3, but with a standard deviation of 100. The utility function for this problem was defined as the inverse of the determinant of the variance of the indirect inference posterior distributions for (μ_L, ν);

$$u(d, \mathbf{y}) = 1/\det(\widehat{\mathrm{Var}}(\mu_L, \nu | \mathbf{y}, d, \boldsymbol{\phi})), \tag{14}$$

where $\boldsymbol{\phi} = \boldsymbol{\phi}(\boldsymbol{\theta}, \mathbf{x})$. Other choices of utility functions could be used such as the inverse of the trace of the covariance matrix or the precision of μ_L or ν.

The optimal designs found by the indirect inference approach were similar to the optimal designs inferred by the ABC approach (Drovandi & Pettitt, 2013). Results are displayed in Table 2. As in the death model example, Gaussian kernel density estimation was used to find the optimal design, with an automatically selected bandwidth, which was tempered by smoothing factors $h \in (0.1, 0.2, ..., 4)$ and restricted from 24 to 300 days.

5 Discussion

A novel approach to Bayesian experimental design has been presented for models with intractable likelihoods. The use of indirect inference and the design algorithm of Müller

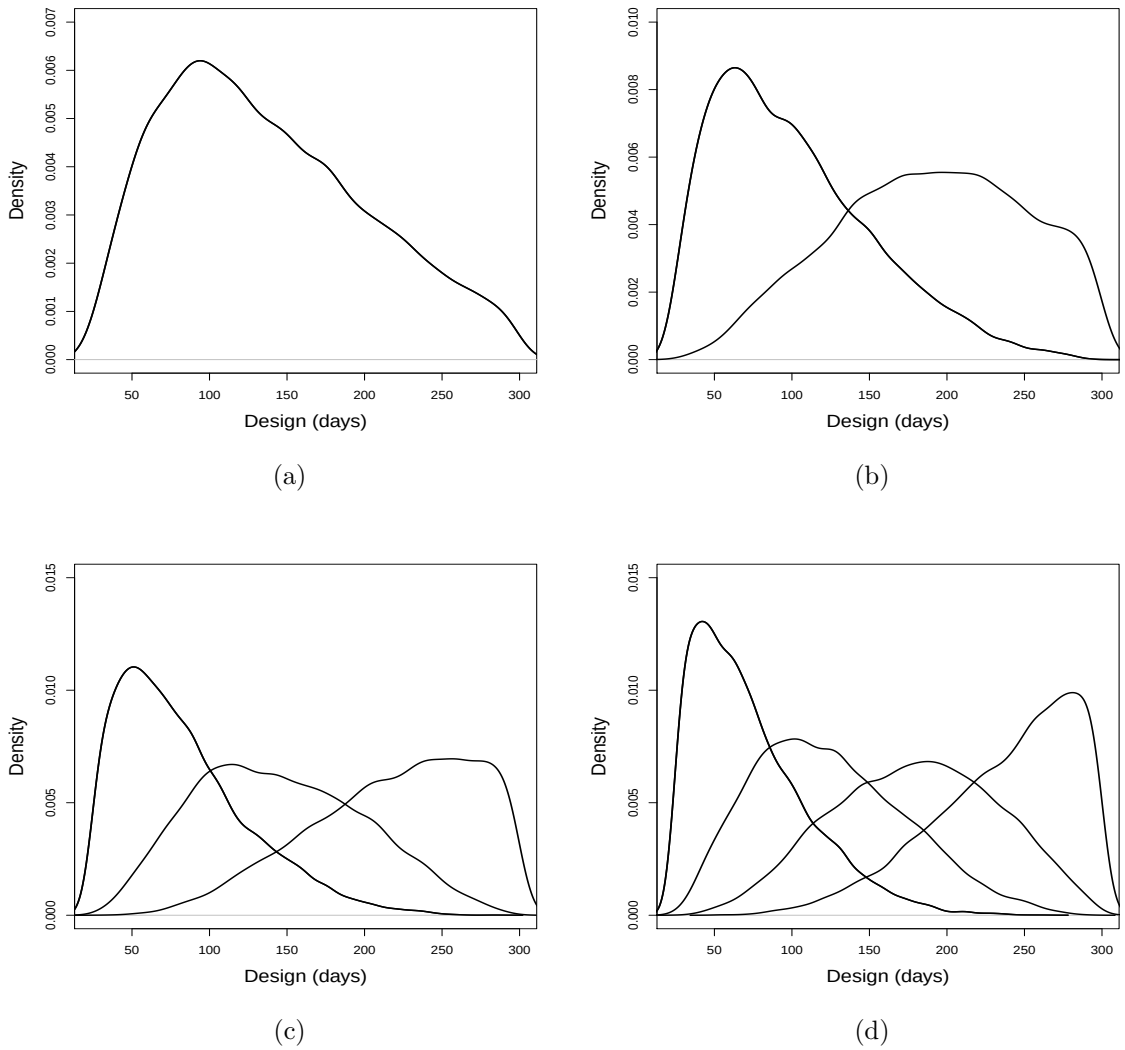

Figure 4: Marginal density estimates of the designs for up to 1 to 4 observations of the macroparasite model (plots (a) to (d) respectively) for $100,000$ iterations of the design algorithm of Müller (1999) using indirect inference.

Table 2: Optimal designs for the macroparasite model for $k = 1, \ldots, 4$ observations, using the design algorithm of Müller (1999) and indirect inference and using ABC (results from Drovandi & Pettitt (2013)). Both approaches use a Gaussian smoothing kernel to find the modal values. Expected utilities, Equation 12, are calculated using indirect inference at the optimal design inferred by each approach.

k	II Optimal design	$u(d)$	ABC Optimal design	$u(d)$
1	93.7	4.971×10^{11}	99	4.971×10^{11}
2	$(66.3, 116.0)$	5.488×10^{11}	$(71, 127)$	5.457×10^{11}
3	$(68.2, 101.3, 156.5)$	5.932×10^{11}	$(95, 105, 231)$	5.841×10^{11}
4	$(63.0, 105.2, 170.7, 238.3)$	6.300×10^{11}	$(79, 121, 231, 273)$	6.262×10^{11}

(1999) allows sampling from an augmented joint probability distribution without calculating the likelihood. The marginal mode of this distribution is the optimal time to observe the stochastic process. The indirect inference approach is a more flexible approach than the ABC approach of Drovandi & Pettitt (2013). It is carried out over a continuous design space, unlike the ABC approach, which is carried out over a discrete set of designs. It avoids the usual deficiencies of ABC such as large storage requirements and the necessity to choose a tolerance level. As a result it could be extended to more complex design problems such as designs with many experimental observations or multi-dimensional designs, which would be difficult in the ABC approach. Furthermore, a new utility was proposed which improved mixing and hence required less computation time and a lower tempering value J of the target distribution.

The methodology was demonstrated using two examples and comparisons were made to published results from existing methods. The death model served to introduce the approach and results obtained were similar to the optimal designs inferred by the moment closure approach (Cook et al., 2008). Marginal densities of the designs were similar to the same inference carried out using the true tractable likelihood. The motivating example was a stochastic epidemiological model for the evolution of macroparasites. The optimal designs found by the indirect inference approach for both examples were similar to other approaches in the literature such as moment closure and ABC. This highlights the success of the indirect inference approach to optimal fully Bayesian experimental design. Furthermore, the moment closure approximation is limited in its application and the ABC approach to optimal Bayesian design has several drawbacks. The indirect inference approach helps us overcome some of these drawbacks.

An important advantage over existing methods of the proposed indirect inference approach to Bayesian experimental design for models with intractable likelihoods is that the indirect inference approach can be extended to include other design variables, without incurring additional parameter storage costs. The ABC approach of Drovandi & Pettitt (2013) suffers from a restriction in the dimensionality of the design space due to storage requirements. For example, in the macroparasite model, the number of initial larvae injected in the

host cats could be incorporated in the inference for optimal design as well as the optimal observation times. The auxiliary model could be altered to reflect this, for example, by expressing the overdispersion as a polynomial function of the initial larvae. One would expect that the maximum number of available larvae would be optimal as this would provide more informative data. But this would come at an experimental cost. A useful example to consider is the optimal design for a clinical trial, where the optimal drug dose and observation times of patient response could be estimated jointly by this approach. The ABC approach would have difficulty storing all of the prior predictive data produced by discretising two design dimensions. In the approach using indirect inference, all that is required is the storage of the look-up table, $\{\boldsymbol{\theta}^i, \boldsymbol{\phi}^i\}_{i=1}^n$, regardless of the design dimension.

One difficulty with this design problem results from the fact that, for more than one observation of the stochastic process, the observations are time ordered. This means that the multi-dimensional density estimates of the design are defined on a simplex. For example, for two ordered sampling times, uniform draws would see marginal modes at the end points despite there being no unique multivariate mode over the simplex. The modes of the marginal densities in Figures 3 and 4 are drawn towards endpoints d_{min} and d_{max}. Kernel density estimation is used to find the true modes. As in Drovandi & Pettitt (2013), we take a somewhat ad-hoc approach to temper the optimal bandwidth selected by the plug-in method of Wand & Jones (1994) by various smoothing factors. The mode corresponding to the highest utility is chosen as the optimal design. Bandwidth selection is fixed across the space, which may not be appropriate. Ideas for further research include ascertaining a more principled approach to this problem such as adaptive kernel density estimation or through the use of Gaussian processes. Alternatively, a principled approach is to re-parameterise a large number of experimental observations times over a lower dimensional space as in Ryan et al. (2014), where the design algorithm of Müller (1999) is operated over the lower dimensional parameterisation using the quantiles of a beta distribution. This is a flexible approach requiring the optimisation of only two parameters of the beta distribution. Using this methodology, the use of indirect inference could be extended to optimise a large number of experimental observations, for example, 10 or more. This would be problematic for the ABC approach due to an increase in tolerance for high dimensional designs, which can only be offset by an increase in the storage requirements.

In the current setup, the offline step requires a training design d_T, which is used to estimate the relationship between the generative model and the auxiliary model through a noisy mapping $\boldsymbol{\phi}(\boldsymbol{\theta}, \mathbf{x})$. For simplicity we used the same structure as a previous experiment. But it may be possible to improve estimation by relaxing this restriction. The choice of designs d_T from which to generate simulations $\mathbf{x}$ is itself a design problem and requires further consideration in future extensions and applications of this approach. Other ideas relating to the offline step include the smoothing of the estimate $\boldsymbol{\phi}(\boldsymbol{\theta}, \mathbf{x})$ of $\boldsymbol{\phi}(\boldsymbol{\theta})$, without incurring the increased computational cost as $m \to \infty$. As discussed in Section 3, under the assumption that the auxiliary estimator is consistent, the true mapping $\boldsymbol{\phi}(\boldsymbol{\theta})$, for a particular value of $\boldsymbol{\theta}$, can be recovered as $m \to \infty$. Possibilities to avoid the computational cost of increasing m and simultaneously (perhaps) improve inference, include the use of a

local multivariate smoother such as a spline, a kernel smoother or a Gaussian process, to recover an estimate $\widehat{\boldsymbol{\phi}}(\boldsymbol{\theta})$ of $\boldsymbol{\phi}(\boldsymbol{\theta})$, based on the noisy mapping $\boldsymbol{\phi}(\boldsymbol{\theta}, \mathbf{x})$.

Acknowledgements

The work of Caitríona M. Ryan was supported by the Australian Research Council Discovery Project DP1210269 and that of Christopher C. Drovandi and Anthony N. Pettitt by an Australian Research Council Discovery Project DP110100159.

References

C. Andrieu & G. O. Roberts (2009). 'The pseudo-marginal approach for efficient Monte Carlo computations'. *The Annals of Statistics* **37**:697–725.

D. A. Berry (2004). 'Bayesian statistics and the efficiency and ethics of clinical trials'. *Statistical Science* **19**(1):175–187.

W. S. Cleveland (1979). 'Robust locally weighted regression and smoothing scatterplots'. *Journal of the American Statistical Association* **74**(368):829–836.

A. R. Cook, et al. (2008). 'Optimal observation times in experimental epidemic processes'. *Biometrics* **64**(3):860–868.

D. A. P. Delzell, et al. (2012). 'Key properties of D-optimal designs for event-related functional MRI experiments with application to nonlinear models'. *Statistics in Medicine* **31**(29):3907–3920.

D. A. Denham, et al. (1972). 'Studies with *Brugia pahangi*.-I. Parasitological observations on primary infections of cats (Felis catus)'. *International Journal for Parasitology* **2**(2):239–247.

C. C. Drovandi, et al. (2013). 'Sequential Monte Carlo for Bayesian sequentially designed experiments for discrete data'. *Computational Statistics & Data Analysis* **57**(1):320–335.

C. C. Drovandi & A. N. Pettitt (2013). 'Bayesian experimental design for models with intractable likelihoods'. *Biometrics* **69**(4):937–948.

C. C. Drovandi, et al. (2011). 'Approximate Bayesian computation using indirect inference'. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **60**(3):317–337.

C. C. Drovandi, et al. (2014). 'Bayesian Indirect Inference using a parametric auxiliary model'. Tech. rep., Queensland University of Technology, http://eprints.qut.edu.au/63767/.

R. A. Fisher (1926). 'The arrangement of field experiments'. *Journal of the Ministry of Agriculture of Great Britain* **33**:503–513.

N. J. Fox, et al. (2013). 'Modelling Parasite Transmission in a grazing system: the importance of host behaviour and immunity'. *PloS one* **8**(11):e77996.

A. R. Gallant & R. E. McCulloch (2009). 'On the determination of general scientific models with application to asset pricing'. *Journal of the American Statistical Association* **104**(485):117–131.

A. R. Gallant & G. Tauchen (1996). 'Which moments to match?'. *Econometric Theory* **12**(04):657–681.

D. T. Gillespie (1977). 'Exact stochastic simulation of coupled chemical reactions'. *The Journal of Physical Chemistry* **81**(25):2340–2361.

C. Gourieroux, et al. (1993). 'Indirect inference'. *Journal of Applied Econometrics* **8**(S1):S85–S118.

X. Huan & Y. Marzouk (2012). 'Gradient-based stochastic optimization methods in Bayesian experimental design'. *International Journal for Uncertainty Quantification* .

C. Kreutz & J. Timmer (2009). 'Systems biology: experimental design'. *FEBS Journal* **276**(4):923–942.

I. Krishnarajah, et al. (2005). 'Novel moment closure approximations in stochastic epidemics'. *Bulletin of Mathematical Biology* **67**(4):855–873.

S. Kullback & R. A. Leibler (1951). 'On information and sufficiency'. *The Annals of Mathematical Statistics* **22**:76–86.

D. V. Lindley (1972). *Bayesian statistics, a review.* Capital City Press, Montepelier, Vermont.

E. Michael, et al. (1998). 'Modelling variability in lymphatic filariasis: Macrofilarial dynamics in the *Brugia pahangi*–cat model'. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **265**(1391):155–165.

P. Müller (1999). 'Simulation-based optimal design'. *Bayesian Statistics* **6**:459–474.

J. A. Nelder & R. Mead (1965). 'A simplex method for function minimization'. *Computer Journal* **7**(4):308–313.

E. A. Ottesen (2006). 'Lymphatic filariasis: treatment, control and elimination'. *Advances in Parasitology* **61**:395–441.

N. Paciorkowski, et al. (2000). 'B1 B lymphocytes play a critical role in host protection against lymphatic filarial parasites'. *The Journal of Experimental Medicine* **191**(4):731–736.

R. W. Reeves & A. N. Pettitt (2005). 'A theoretical framework for approximate Bayesian computation'. In *Proceedings of the 20th International Workshop on Statistical Modelling, Sydney*, pp. 393–396.

E. Renshaw (1991). *Modelling biological populations in space and time*. Cambridge University Press, Cambridge.

S. Riley, et al. (2003). 'Robust parameter estimation techniques for stochastic within-host macroparasite models'. *Journal of Theoretical Biology* **225**(4):419–430.

E. G. Ryan, et al. (2014). 'Towards Bayesian experimental design for nonlinear models that require a large number of sampling times'. *Computational Statistics & Data Analysis* **70**:45–60.

A. A. Smith (1993). 'Estimating nonlinear time-series models using simulated vector autoregressions'. *Journal of Applied Econometrics* **8**(S1):S63–S84.

R. R. Suswillo, et al. (1982). 'The number and distribution of *Brugia pahangi* in cats at different times after a primary infection.'. *Acta Tropica* **39**(2):151–156.

P. J. M. Van Laarhoven & E. H. L. Aarts (1987). *Simulated annealing: Theory and Applications*. D. Reidel Publishing Company, Dordrecht.

M. P. Wand & M. C. Jones (1994). 'Multivariate plug-in bandwidth selection'. *Computational Statistics* **9**(2):97–116.